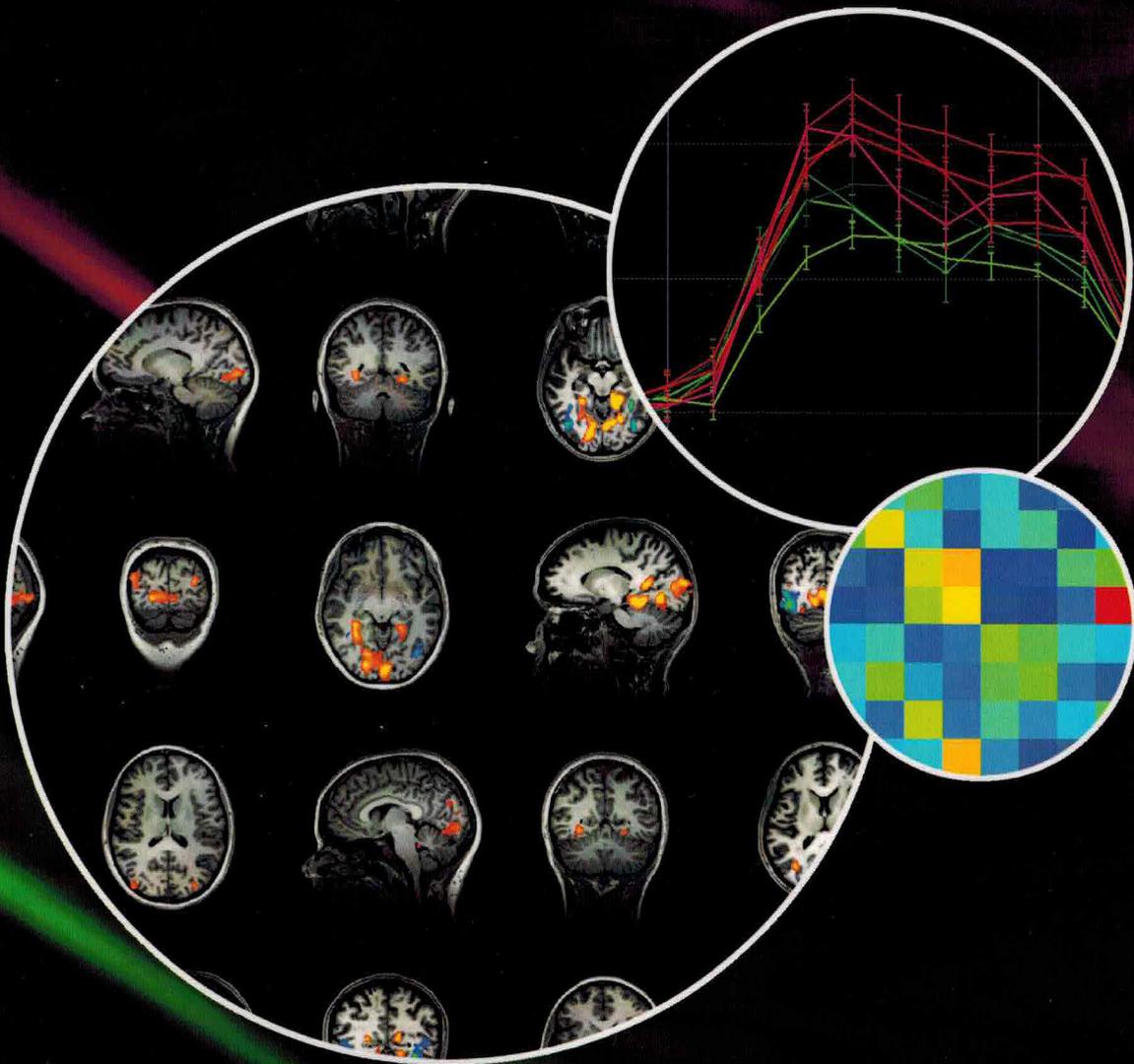


ISSN 1053-8119  
Volume 125, January 15, 2016

# NeuroImage

*Editor-in-Chief*  
**Peter Bandettini**



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

**ScienceDirect**



## Feature diagnosticity and task context shape activity in human scene-selective cortex



Matthew X. Lowe<sup>a,b,\*</sup>, Jason P. Gallivan<sup>c,d</sup>, Susanne Ferber<sup>b,e</sup>, Jonathan S. Cant<sup>a</sup>

<sup>a</sup> Department of Psychology at Scarborough, University of Toronto, Toronto, ON M1C 1A4, Canada

<sup>b</sup> Department of Psychology at St. George, University of Toronto, Toronto, ON M5S 3G3, Canada

<sup>c</sup> Centre for Neuroscience Studies, Queen's University, Kingston, ON K7L 3N6, Canada

<sup>d</sup> Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada

<sup>e</sup> Rotman Research Institute, Baycrest, Toronto, ON M6A 2E1, Canada

### ARTICLE INFO

#### Article history:

Received 8 July 2015

Accepted 31 October 2015

Available online 2 November 2015

#### Keywords:

Scene recognition

Texture

Layout

Ventral visual cortex

Feature integration

Attention

### ABSTRACT

Scenes are constructed from multiple visual features, yet previous research investigating scene processing has often focused on the contributions of single features in isolation. In the real world, features rarely exist independently of one another and likely converge to inform scene identity in unique ways. Here, we utilize fMRI and pattern classification techniques to examine the interactions between task context (i.e., attend to diagnostic global scene features; texture or layout) and high-level scene attributes (content and spatial boundary) to test the novel hypothesis that scene-selective cortex represents multiple visual features, the importance of which varies according to their diagnostic relevance across scene categories and task demands. Our results show for the first time that scene representations are driven by interactions between multiple visual features and high-level scene attributes. Specifically, univariate analysis of scene-selective cortex revealed that task context and feature diagnosticity shape activity differentially across scene categories. Examination using multivariate decoding methods revealed results consistent with univariate findings, but also evidence for an interaction between high-level scene attributes and diagnostic visual features within scene categories. Critically, these findings suggest visual feature representations are not distributed uniformly across scene categories but are shaped by task context and feature diagnosticity. Thus, we propose that scene-selective cortex constructs a flexible representation of the environment by integrating multiple diagnostically relevant visual features, the nature of which varies according to the particular scene being perceived and the goals of the observer.

© 2015 Elsevier Inc. All rights reserved.

### 1. Introduction

How does the brain process the environment around us? Since the initial description of the scene-selective parahippocampal place area (PPA; Epstein and Kanwisher, 1998), investigations have sought to answer this question by attempting to clarify the nature of the neural representations in this region. Much of this research has revealed a primary role for PPA in the encoding of spatial features within a scene, such as structural geometry or layout (Epstein and Kanwisher, 1998; Epstein et al., 2003), spatial boundary (Park et al., 2011), and spatial depth (Kravitz et al., 2011). Conversely, recent studies support the notion that its neural representations extend beyond spatial features and include the encoding of non-spatial contextual associations of objects (Bar et al., 2008), high-level conceptual scene categories (Walther et al., 2009; 2011; Dilks et al., 2011), and surface texture and material properties (Peuskens et al., 2004; Cant and Goodale, 2007; 2011). In order to better understand

scene representation, however, it is not only necessary to understand the contributions of individual features, but also how these features converge to contribute to the formation of scene identity. Yet disentangling feature-specific modulation of scene-selective neural activity within global scene representations remains a challenge, as these features rarely exist in isolation, and may inform scene identity through complex interactions which vary according to scene category.

Early research exploring diagnostic visual features in the recognition of objects revealed a primary role for edge-based information (i.e., structure), suggesting surface characteristics such as color and texture play only a secondary role in object recognition (Biederman and Ju, 1988). Research has since extended support for edge-based determinants of visual object recognition to scene perception (Delorme et al., 2000; Walther et al., 2011; Walther and Shen, 2014), yet a growing body of work suggests diagnostic surface characteristics such as color and texture are instrumental in mediating early-stage scene gist processing that is responsible for successful scene recognition (Schyns and Oliva, 1994; Oliva and Schyns, 2000; Goffaux et al., 2005; Steeves et al., 2004; Castelano and Henderson, 2008). Given these differences, the interplay between surface properties and structural features as determinants for

\* Corresponding author at: Department of Psychology, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.

E-mail address: [matthew.lowe@mail.utoronto.ca](mailto:matthew.lowe@mail.utoronto.ca) (M.X. Lowe).

scene recognition is currently unclear. One framework for scene perception, which may reconcile these differences, proposes that the recognition of complex visual scenes can be understood through interactions between perceptually available information and categorization demands (Oliva and Schyns, 1997). This recognition framework centers on the notion of feature diagnosticity: the idea that specific visual cues are used for specific types of categorizations and an interaction between task demands and available visual information can explain how different cues are used to recognize scenes. In other words, diagnostic visual features may emerge as a function of their usefulness in defining the identity of a scene, and the task demands placed on the observer. Thus, the present study aims to investigate the influence of diagnostic surface- and edge-based visual features on neural scene processing across a range of scene categories.

Given the variability of visual information across scene categories, Oliva and Torralba (2006) proposed that the most effective global features for scene identification will be those capturing the global structure and meaning of the visual world. For example, manufactured environments (e.g., cities) are dominated by prominent edge-based information containing straight horizontal and vertical lines, while natural landscapes (e.g., deserts) tend to have zones of characteristic textures and undulating contours which may be meaningful for scene identification (Oliva and Torralba, 2001). Thus, structural information (e.g., layout and geometry) may be of greater diagnostic relevance when discriminating scenes within manufactured environments, whereas both distinctive textured zones and undulating spatial structures may be diagnostic for scene identification in natural environments. Indeed, behavioral research has revealed the importance of global texture cues in capturing the diagnostic structure of natural scenes (Oliva and Torralba, 2006). For instance, a forest can be described in terms of the roughness and homogeneity of its textural components, providing meaningful information to a human observer comparing two forest scenes (Rao and Lohse, 1993). The neural representations of texture perception in PPA, however, have been investigated using isolated objects, and not entire scenes (Cant and Goodale, 2007; 2011), and it is therefore unclear how texture contributes to scene representations in scene-selective cortex.

In light of the importance of layout and texture information in scene perception, and potential differences in the relevance of these features for categorizing different scenes, the present study examined neural activity in scene-selective cortex while observers attended to either the layout or texture of natural and manufactured scenes, either of which could change while the other was held constant. We hypothesized that PPA would show equal sensitivity to manipulations of both layout and texture in natural scenes, where textured zones and layout may be equally relevant for distinguishing scene identity. In manufactured scenes, however, we hypothesized that PPA would show less sensitivity to texture, relative to layout, as these scenes contain prominent horizontal and vertical structural components that can aid in the discrimination of scene identity. In order to isolate effects to PPA, we also examined the modulation of brain activity in other areas of scene- and object-processing networks, and additionally localized a region of early visual cortex to examine if activation patterns observed in PPA can be dissociated from activity in early visual areas. Building on previous research (Walther et al., 2009; 2011; Park et al., 2011; Kravitz et al., 2011), we took advantage of both univariate and multivariate analyses to investigate previously unexplored questions of how task-dependent global scene features (i.e., attend to texture or layout) interact with high-level conceptual scene attributes (i.e., content: natural vs. manufactured scenes; and spatial boundary: open vs. closed scenes) to shape scene representation in human visual cortex.

## 2. Materials and methods

### 2.1. Observers

Twelve paid observers (6 males; mean age  $27.4 \pm 3.8$  years) with normal or corrected-to-normal visual acuity were recruited from the

University of Toronto community. Observers gave informed consent in accordance with the University of Toronto Ethics Review Board.

### 2.2. Stimuli and procedure

Stimuli were grayscale photographs from four different scene categories devoid of foreground objects to avoid interference (see Davenport and Potter, 2004; Joubert et al., 2007) and created by varying features of spatial boundary (open vs. closed) and scene content (natural vs. manufactured; Fig. 1) (Oliva and Torralba, 2001). After selection of our four scene categories, twelve unique structural arrangements (i.e., layouts) were selected for each category, and twelve appropriate textures were applied to the dominant surface of each layout (mapped onto scene gradient and depth using Adobe Photoshop CS3), yielding 144 unique images per scene category (12 layouts/category  $\times$  12 textures/layout  $\times$  4 scene categories = 576 total images). E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to control stimulus presentation and collect behavioral responses. Images were rear-projected onto a screen in the MRI scanner at a resolution of  $500 \times 500$  pixels (subtending  $10.4^\circ \times 10.4^\circ$  of visual angle), and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes. We used a blocked fMRI experimental paradigm, wherein sixteen images from a single scene category were presented in blocks of 16-s each. Each block was preceded by a 12-s fixation period and a 4-s written instruction to attend to changes in either the texture or layout of the scenes in the ensuing block.

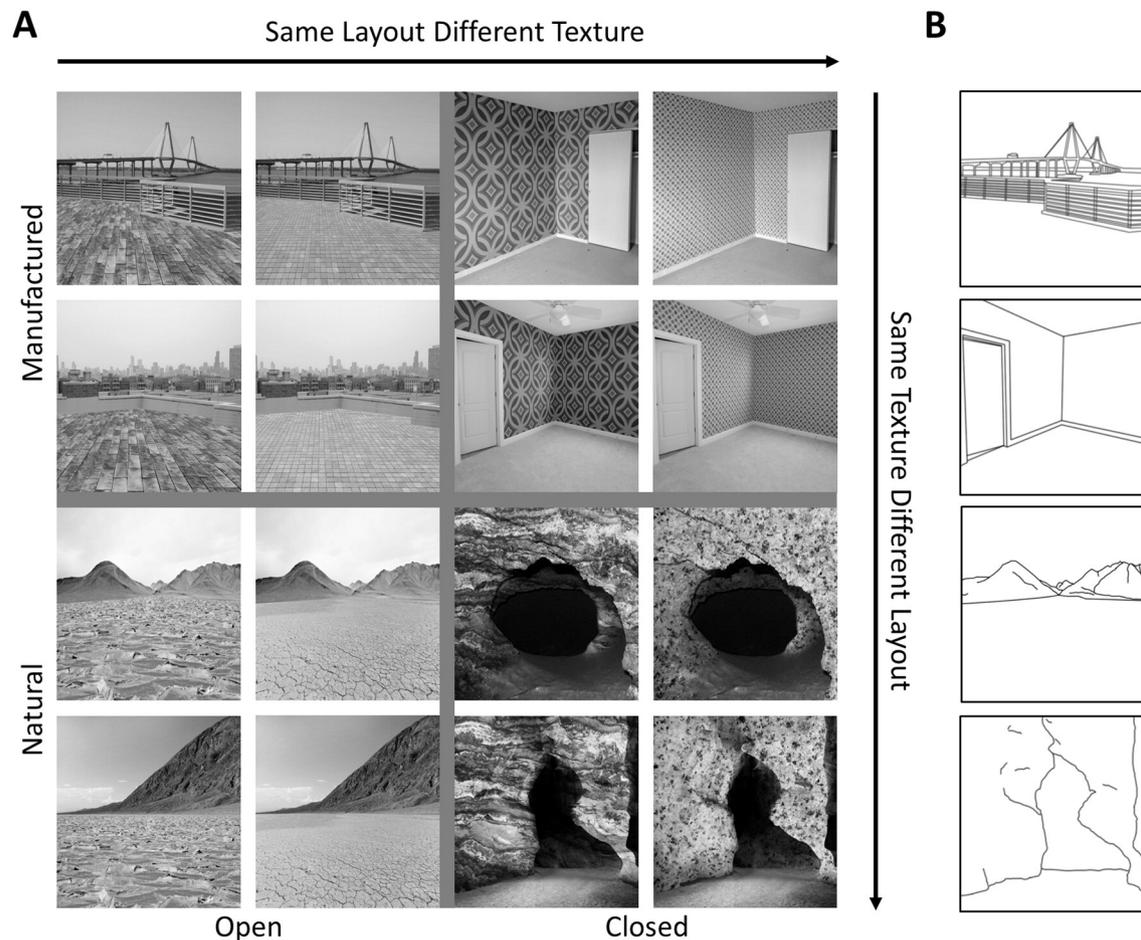
In each trial (8 per block, lasting 2 s each), two scenes were presented for 300 ms (separated by a 200-ms blank interval), and the task of the observers was to decide if the attended feature (i.e., layout or texture) was the same or different across the two images, responding during a 1.5-s period following the onset of the second image (via a response pad placed in the observer's right hand). Each block contained an equal number of "same" and "different" trials. Observers were instructed to maintain central fixation and respond as accurately as possible, placing no emphasis on fast response times to help encourage accurate performance. Images from a single scene category were presented randomly within each block, and each image could be repeated only once per observer. Each observer took part in 8 runs (4 min 28 s each). Each run contained a unique and counterbalanced order of 8 different stimulus blocks (i.e., 8 different conditions: attend to texture or layout in each of the four scene categories). Run order was randomized across observers, and scene category was held constant per block.

### 2.3. Localizer scan

Stimuli used to localize object-, scene-, and face-sensitive areas of cortex, as well as early visual cortex, were photographs of various scenes, faces, common objects, and tile-scrambled images. Stimuli were presented in 16-s blocks of 32 images at a resolution of  $375 \times 375$  pixels ( $7.8^\circ \times 7.8^\circ$ ) and were displayed for 400 ms each, with an interstimulus interval of 50 ms. Observers fixated on a centrally presented black fixation cross and were instructed to respond with a button press when the fixation cross changed from black to red (randomly occurring once or twice per stimulus block). There were 4 blocks for each stimulus category within a run, and there were two unique run orders. Each observer took part in three localizer runs (6 min 40 s each).

#### 2.3.1. MRI acquisition

Scanning was performed at the Center for Addiction and Mental Health using a 3-T GE Discovery MR750 whole-body MRI scanner equipped with an 8-channel head coil. T1-weighted anatomical images were acquired using a 3D SAG T1 BRAVO spiral pulse sequence [repetition time (TR), 6736 ms; echo time (TE), 3 ms; inversion time, 650 ms; flip angle  $8^\circ$ ,  $256 \times 256$  matrix size, 200 slices, 1 mm isovoxel]. For the functional runs, T2\*-weighted images sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a spiral pulse



**Fig. 1.** Experimental stimuli. (A) Examples of the four scene categories used. Scenes were defined by their spatial boundary (open vs. closed) and content (natural vs. manufactured). Observers attended to either the global texture or spatial layout (task context) within a scene, either of which could change while the other was held constant. (B) Comparisons of structural information displayed through line drawings. Note that, in the absence of textural information, structure is more informative in manufactured scenes.

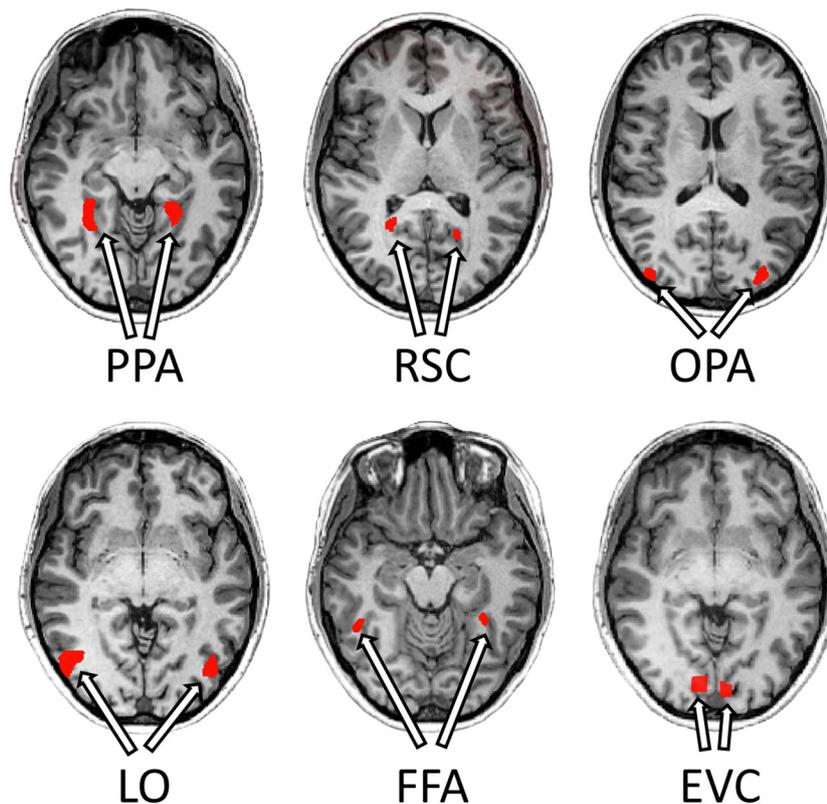
sequence ( $64 \times 64$  matrix size; field of view 22 cm; TR 2000 ms; TE 30 ms; flip angle  $60^\circ$ ; 200 volumes for the localizer runs and 134 volumes for the main experimental runs). Thirty-one slices ( $3.4 \text{ mm} \times 3.4 \text{ mm} \times 5 \text{ mm}$ , no gap) parallel to the anterior and posterior commissure line were collected in all functional runs.

#### 2.4. Univariate data analysis

fMRI data were processed and analyzed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, the Netherlands). Data preprocessing included slice acquisition time correction, 3D motion correction, temporal filtering (linear trend removal and high-pass filtering set at 3 cycles/run), and Talairach space transformation (Talairach and Tournoux, 1988). Data from the functional localizer was analyzed using a general linear model (GLM), accounting for hemodynamic response lag (Friston et al., 1994). Regions of interest (ROIs) can be seen in Fig. 2. In accordance with Epstein and Kanwisher (1998), the PPA ROI was defined as a region in the collateral sulcus and parahippocampal gyrus whose activation was higher for scenes than for faces and objects (false discovery rate,  $q < 0.05$ ; this threshold applies to all functional regions localized in individual observers; identified in all twelve observers; see Fig. 2). In addition, in accordance with Epstein and Higgins (2007) and Dilks et al. (2013), retrosplenial complex (RSC) and the occipital place area (OPA; also known as transverse occipital sulcus) ROIs were functionally defined as regions in retrosplenial cortex–posterior cingulate–medial parietal cortex and

transverse occipital cortex, respectively, whose activations were higher for scenes than for faces and objects (identified in eleven, and nine, observers, respectively). In accordance with Grill-Spector et al. (2000), the lateral occipital area (LO) was defined as a region in the lateral occipital cortex near the posterior inferotemporal sulcus, with activation higher for objects than for tile-scrambled objects (identified in all twelve observers). Early visual cortex (EVC) was defined as a retinotopic region around the calcarine sulcus with activation higher for scrambled objects than intact objects (MacEvoy and Epstein, 2011; Cant and Xu, In Press; identified in eleven observers).

Following the standard ROI-based analysis approach (Saxe et al., 2006), we overlaid the ROIs onto the data from our main experiment and extracted time courses from each observer. Peak responses for each condition were obtained by collapsing the time courses for all of the conditions and then identifying the time point of greatest signal amplitude in the average response (Xu and Chun, 2006; Xu, 2010; Cant and Xu, 2012). This was done separately for each observer in each ROI, and the resultant peak responses were then averaged across all observers. The average levels of peak activation (measured in percent BOLD signal change from baseline fixation) for each condition across observers were subjected to a  $6$  (ROI: PPA, RSC, OPA, LO, FFA, EVC)  $\times 2$  (spatial boundary: open vs. closed)  $\times 2$  (content: natural vs. manufactured)  $\times 2$  (task: texture vs. layout) repeated-measures ANOVA (SPSS, Chicago, IL, USA). This analysis revealed differences in activation across ROIs, so further analyses were conducted on each ROI separately. Moreover, subsequent analyses revealed no significant differences in activation between open



**Fig. 2.** Regions of interest. Functionally defined ROIs are shown on a representative participant's brain. Talairach coordinates for peak voxels of each ROI in this representative participant are shown as follows: LPPA,  $-22, -43, -9$ ; RPPA,  $31, -45, -5$ ; LRSC,  $-16, -54, 7$ ; RRSC,  $22, -49, 7$ ; LOPA,  $-29, -84, 15$ ; ROPA,  $40, -83, 17$ ; LLO,  $-39, -72, -7$ ; RLO:  $45, -69, 2$ ; LFFA,  $-36, -34, -23$ ; RFFA,  $45, -45, -12$ ; LEVC,  $-6, -87, -3$ ; REVC,  $10, -84, -4$ .

and closed scenes in scene-selective cortex, so all subsequent univariate analyses were conducted by examining differences in content and task, collapsed across spatial boundary (see univariate results below). Planned pairwise comparisons (Bonferroni-corrected for multiple comparisons) were then conducted to examine the relationship in activity between texture and layout for natural and manufactured scenes in each ROI. Since no differences between hemispheres were observed for all ROIs, bilateral regions were combined for analysis.

## 2.5. Multivoxel pattern analysis (MVPA)

### 2.5.1. Support vector machine classifiers

Pattern classification was performed with a combination of in-house software (using Matlab) and the Princeton MVPA Toolbox for Matlab (<http://code.google.com/p/princeton-mvpa-toolbox/>) using a Support Vector Machines (SVM) classifier (libSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). The SVM model used a linear kernel function and a constant cost parameter,  $C = 1$ , to compute a hyperplane that best separated the block/condition responses. To test the accuracy of the SVM classifiers, we used a "leave-one-run-out" N-fold cross-validation, in which a single fMRI run was reserved for classifier testing. We performed this N-1 cross-validation procedure until all runs were separately tested, and then averaged across N-iterations in order to produce a representative classification accuracy measure for each participant, ROI, and pattern discrimination (see Duda et al., 2001).

### 2.5.2. Multiclass and pairwise discriminations

SVMs are designed for classifying differences between two stimuli and LibSVM (the SVM package implemented here) uses the so-called "one-against-one method" for classification (Hsu and Lin, 2002). With

the SVMs, we performed two complementary types of classification analyses: one in which the multiple pairwise results were combined in order to produce multiclass discriminations (distinguishing among all 8 of our condition types) and another in which the individual pairwise discriminations were examined and tested separately.

The multiclass discrimination approach allowed for an examination of the distribution of the classifier guesses through visualization of the resulting "confusion matrix." In a confusion matrix, each row ( $i$ ) represents the instances of the actual condition and each column ( $j$ ) represents the predicted condition. Their intersection ( $i, j$ ) represents the (normalized) number of times a given condition  $i$  is predicted by the classifier to be condition  $j$ . Thus, the confusion matrix provides a direct visualization of the extent to which a decoding algorithm confuses (or correctly identifies) the different classes. All correct classifications are located in the diagonal of the matrix (with classification errors represented by non-zero values outside of the diagonal) and average decoding performance is defined as the mean across the diagonal. The values in each row sum to 1 (i.e., 100% classification). If decoding is at chance levels, then classification performance will be at  $1/8 = 12.5\%$ . For all multiclass discriminations, we statistically assessed decoding significance across participants (for each ROI and condition epoch) using one-tailed  $t$ -tests versus 12.5% chance decoding.

In contrast, the pairwise discrimination approach allowed us to identify ROIs encoding scene content and spatial boundary while other scene attributes were held constant, as well as make comparisons across tasks (i.e., texture and layout). It is important to recognize that this pairwise information and any nuances in the pattern of effects would be largely obscured using a multiclass discrimination approach. For pairwise discriminations, we statistically assessed decoding significance across participants using one-tailed  $t$ -tests versus 50% chance

decoding. Importantly, for both the multiclass and pairwise discriminations, an FDR correction of  $q \leq 0.05$  was applied based on the number of ROIs examined (Benjamini and Hochberg, 1995).

2.5.3. Inputs to the SVM Classifier

BOLD percent signal change values for each ROI provided inputs to the SVM classifier. The percent signal change response was computed from the time-course activity for the task-evoked responses with respect to the time-course of a run-based averaged baseline fixation value, for all voxels in the ROI. The baseline fixation window was defined as a time point prior to the 4-s instruction period before each stimulus block (6 s prior to block onset, averaged across all blocks within an experimental run). For the block-evoked activity, we extracted, for each condition, the average of imaging volumes 3–8 (i.e., 6–16 s), which are time points encompassing the first peak of the hemodynamic response until the end of the experimental block. This windowed-average percent signal change classification approach with that used in recent work using the same technique (e.g., Gallivan et al., 2013; Gallivan et al., 2014). Following the extraction of each block's activity, these values were rescaled between  $-1$  and  $+1$  for each voxel pattern within an ROI (Misaki et al., 2010).

2.6. Behavioral data analysis

Behavioral performance measures of accuracy were recorded using E-Prime 2.0 software and analyzed with SPSS, by performing a 2 (spatial boundary: open vs. closed)  $\times$  2 (content: natural vs. manufactured)  $\times$  2 (task: texture vs. layout) repeated-measures ANOVA.

3. Results

3.1. Univariate analysis

An initial repeated-measures ANOVA including ROI as a factor (PPA, RSC, OPA, LO, FFA, and EVC) revealed significant main effects of ROI ( $F_{5,30} = 21.01, p < 0.001$ ), content ( $F_{1,6} = 10.43, p = 0.018$ ), and task ( $F_{1,6} = 133.57, p < 0.001$ ), but not spatial boundary ( $F_{1,6} = 0.22, p = 0.66$ ). Additionally, we observed significant interactions between ROI and content ( $F_{5,30} = 12.30, p < 0.001$ ) and task ( $F_{5,30} = 4.95, p = 0.002$ ), demonstrating functional differences in scene-, object-, face-

selective, and early visual cortex for our task and stimuli. Thus, we conducted all subsequent analyses on each ROI individually. Moreover, since no main effect of spatial boundary was observed (and no two-way interactions with this factor and ROI), we collapsed across spatial boundary in subsequent analyses to examine differences between content and task (Fig. 3).

Further examination of scene-selective cortex revealed a significant main effect of content (PPA:  $F_{1,11} = 58.20, p < 0.001$ ; RSC:  $F_{1,10} = 44.29, p < 0.001$ ; OPA:  $F_{1,8} = 6.92, p = 0.030$ ) and task (PPA:  $F_{1,11} = 13.62, p = 0.004$ ; RSC:  $F_{1,10} = 8.73, p = 0.014$ ; OPA:  $F_{1,8} = 11.43, p = 0.010$ ), reflecting higher activity when observers attended to manufactured, over natural, scenes, and layout over texture, respectively. A significant interaction between content and task was found in PPA ( $F_{1,11} = 8.55, p = 0.014$ ) and RSC ( $F_{1,10} = 9.19, p = 0.013$ ), but not OPA ( $F_{1,8} = 3.43, p = 0.101$ ). When we examined object-selective LO, a significant main effect of task ( $F_{1,11} = 9.03, p = 0.012$ ), but not content ( $F_{1,11} = 0.05, p = 0.82$ ) was observed (and no significant interaction), revealing higher activity when observers attended to layout over texture, with no difference between manufactured and natural scenes. In contrast, when we examined activity in both EVC and FFA, no significant main effects or interactions were observed (all  $F_s < 2.39$ ; all  $p_s > 0.15$ ), indicating no differentiation of scene content or task in these ROIs.

Profiles of neural activity for each task (i.e., texture and layout) were consistent among scene-selective regions, revealing no significant region-by-task interactions (all  $F_s < .978$ ; all  $p_s > 0.35$ ). Interestingly, a region-by-content interaction was observed between PPA and OPA ( $F_{1,8} = 7.09, p = 0.029$ ), but not RSC and OPA ( $F_{1,7} = 2.89, p = 0.133$ ). Conversely, profiles of activation in scene-selective cortex were significantly different from those observed in object-selective cortex (significant region-by-content interaction between LO and PPA, RSC, and OPA; all  $F_s > 13.20$ ; all  $p_s < 0.01$ ), face-selective cortex (significant region-by-content interaction between FFA and PPA, RSC, and OPA; all  $F_s > 15.41$ ; all  $p_s < 0.01$ ), and EVC (significant region-by-content interaction between EVC and PPA, and RSC; all  $F_s > 22.76$ ; all  $p_s < 0.002$ ; significant region-by-task interaction between EVC and PPA, and OPA; all  $F_s > 8.99$ ; all  $p_s < 0.020$ ). This demonstrates that our results are distinct to high-level scene-selective visual cortex, and are not likely explained by appealing to differences in low-level visual features across scene categories.

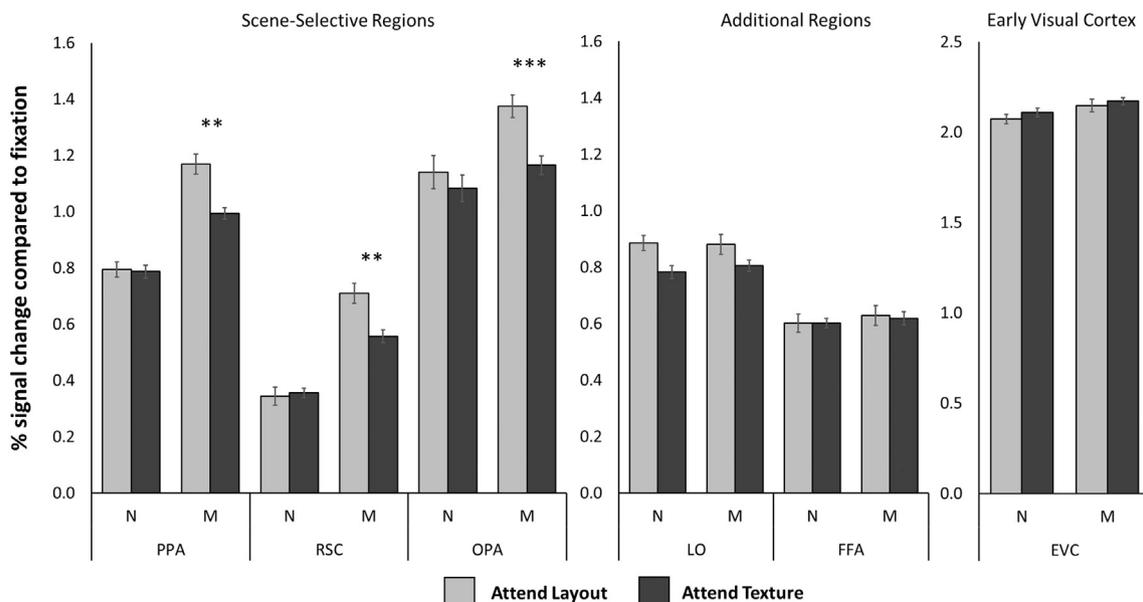


Fig. 3. Univariate results. BOLD signal activation for natural (N) and manufactured (M) scenes when attending to either layout or texture. \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

To test the hypothesis that neural activity in scene-selective cortex would be modulated by the diagnostic relevance of scene features, planned pairwise comparisons were conducted to examine the relationship between scene content (natural vs. manufactured) and task (texture vs. layout) (Fig. 3). In line with our predictions, the results demonstrated equal sensitivity to texture and layout in natural scenes (PPA:  $t_{11} = 0.20$ ,  $p = 0.84$ , RSC:  $t_{10} = 0.43$ ,  $p = 0.67$ ; OPA:  $t_8 = 0.77$ ,  $p = 0.47$ ), but less sensitivity to texture in manufactured scenes (PPA:  $t_{11} = 4.08$ ,  $p = 0.002$ , RSC:  $t_{10} = 3.63$ ,  $p = 0.005$ ; OPA:  $t_8 = 7.29$ ,  $p < 0.001$ ), revealing that the importance of specific scene features (i.e., texture or layout) varies according to the perceived scene content (i.e., natural or manufactured). Notably, no differences were observed in LO, FFA, and EVC (all  $t_s < 1.98$ ; all  $p_s > 0.07$ ).

### 3.2. Multivoxel pattern analysis

Previous research has demonstrated that both the content and spatial boundary of a scene can be decoded from scene-selective regions of cortex (e.g., Park et al., 2011) yet less is known about the neural mechanisms underlying these processes. While there is support that edge-based structural information is sufficient for decoding high-level scene content (Walther et al., 2011), the contributions of edge-based information in defining the spatial boundary of a scene, or how surface characteristics contribute to either of these scene attributes, is unclear. To address these questions, we conducted a number of different multivariate analyses. Following previous investigations (Walther et al., 2009; Park et al., 2011; Kravitz et al., 2011), we first extracted multivoxel fMRI activity and used linear SVM classifiers in each region to examine the extent to which each scene condition could be decoded (Fig. 4A). These multiclass discriminations revealed classification accuracies that were significantly above chance (12.5%) for all ROIs (all  $t_s > 3.39$ , all  $p_s < 0.01$ ), replicating previous findings (see Walther et al., 2009; 2011; Park et al., 2011). To investigate these significant multiclass discriminations in greater detail, we next conducted subsequent analyses to examine interactions of task context with high-level scene attributes within each scene category. We also investigated the classification of task context itself and the distribution of classifier confusion errors across regions of scene-selective cortex. We describe the results of each analysis in turn below.

#### 3.2.1. Classification of high-level scene attributes

While multiclass discriminations allow us to investigate decoding of individual scene categories, it does not provide information about the nature of high-level scene attributes *within* a particular scene category. Thus, as a next step in our multivariate analyses, we conducted pairwise discriminations to examine classification accuracy when decoding scene content (train the classifier on the difference between natural vs. manufactured scenes, and test on the same difference, separately in open and closed scenes, i.e., with spatial boundary held constant; Fig. 4B, D) and spatial boundary (with scene content held constant; Fig. 4C, E) within a given scene category, separately when observers attended to either layout or texture. Examination of scene content while spatial boundary was held constant revealed classification accuracy significantly above chance in PPA for content in both open and closed scenes when observers attended to both layout (Open:  $t_{11} = 4.89$ ,  $p < 0.001$ ; Closed:  $t_{11} = 3.99$ ,  $p = 0.002$ ; Fig. 4B) and texture (Open:  $t_{11} = 2.29$ ,  $p = 0.043$ ; Closed:  $t_{11} = 5.62$ ,  $p = 0.002$ ; Fig. 4D). Similar results were found in OPA for both the attend layout condition (Open:  $t_8 = 5.27$ ,  $p < 0.001$ ; Closed:  $t_8 = 4.11$ ,  $p = 0.003$ ) and the attend texture condition (Open:  $t_8 = 4.97$ ,  $p = 0.001$ ; Closed:  $t_8 = 2.99$ ,  $p = 0.017$ ), and RSC for both the attend layout condition (but note that while RSC showed significant decoding of content in closed scenes,  $t_{10} = 3.91$ ,  $p = 0.003$ , this result was only marginally significant in open scenes,  $t_{10} = 1.98$ ,  $p = 0.076$ ) and the attend texture condition (Open:  $t_{10} = 2.28$ ,  $p = 0.046$ ; Closed:  $t_{10} = 5.38$ ,  $p = 0.003$ ). While classification accuracy in EVC was significantly above chance for scene

content in both open and closed scenes in the layout condition (Open:  $t_{10} = 2.93$ ,  $p = 0.015$ ; Closed:  $t_{10} = 3.78$ ,  $p = 0.004$ ), it was significant in only closed scenes in the texture condition (Open:  $t_{10} = 1.47$ ,  $p = 0.17$ ; Closed:  $t_{10} = 4.02$ ,  $p = 0.002$ ), further dissociating activity in areas of early visual cortex with scene-selective regions.

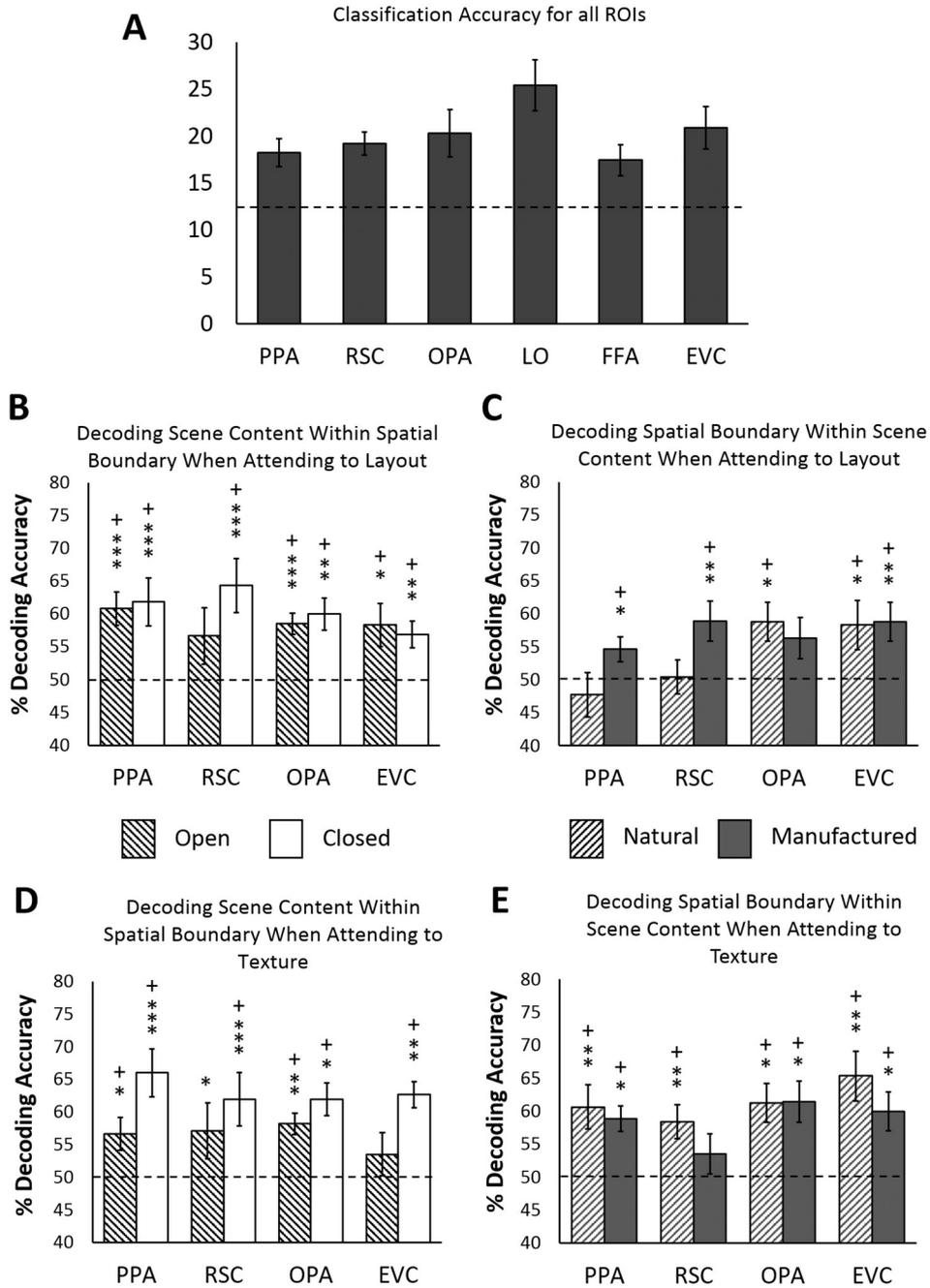
Following previous research using similar stimulus sets (Park et al., 2011; Kravitz et al., 2011), we next examined classification accuracy of spatial boundary (train and test on open vs. closed scenes) while scene content was held constant (either within natural or within manufactured scenes). Specifically, we examined whether we could significantly decode spatial scene features from scene-selective cortex in manufactured, but not natural, scenes, when observers attended to layout, which would be consistent with our initial hypothesis and univariate results. Indeed, the classification of spatial boundary in PPA and RSC was significantly above chance for manufactured (PPA:  $t_{11} = 2.43$ ,  $p = 0.034$ ; RSC:  $t_{10} = 3.35$ ,  $p = 0.007$ ), but not natural, (PPA:  $t_{11} = 1.17$ ,  $p = 0.27$ ; RSC:  $t_{10} = 0.81$ ,  $p = 0.44$ ; Fig. 4C), scenes. Patterns of classification in OPA (Natural:  $t_8 = 3.17$ ,  $p = 0.013$ ; Manufactured:  $t_8 = 2.26$ ,  $p = 0.054$ ), however, were more similar to EVC (Natural:  $t_{10} = 2.61$ ,  $p = 0.026$ ; Manufactured:  $t_{10} = 3.40$ ,  $p = 0.007$ ). Interestingly, unlike the results observed for scene content, when classifying spatial boundary, we observed quite different results when participants attended to the texture, compared with the layout, of a scene (i.e., compare Figs. 4B vs. D with Figs. 4C vs. E). Specifically, these results showed significantly above-chance classification accuracy of spatial boundary in PPA, OPA, and EVC for both natural (PPA:  $t_{11} = 3.91$ ,  $p = 0.002$ ; OPA:  $t_8 = 3.15$ ,  $p = 0.014$ ; EVC:  $t_{10} = 4.17$ ,  $p = 0.002$ ) and manufactured (PPA:  $t_{11} = 2.87$ ,  $p = 0.015$ ; OPA:  $t_8 = 3.27$ ,  $p = 0.011$ ; EVC:  $t_{10} = 2.74$ ,  $p = 0.021$ ) scenes, but only for natural scenes in RSC (natural:  $t_{10} = 3.22$ ,  $p = 0.009$ ; manufactured:  $t_{10} = 1.81$ ,  $p = 0.10$ ).

#### 3.2.2. Classification of task context

As previous reports have indicated sensitivity to processing both texture (e.g., Cant and Goodale, 2007; 2011) and spatial layout (e.g., Epstein and Kanwisher, 1998; Epstein et al., 2003) information in scene-selective cortex, we next examined whether task (texture vs. layout) could be decoded from areas of scene-selective cortex (PPA, RSC, OPA) across each of our four scene categories (see Fig. 1), which may ultimately speak to whether their processing is mediated by shared or distinct neural mechanisms. Critically, no significant decoding was found between these tasks in both PPA and RSC (all  $t_s < 2.44$ , all  $p_s > 0.13$ ), suggesting similar underlying neural representations between the processing of scene texture and layout in these regions (i.e., the patterns of activation for these two attended scene features were quite similar in PPA and RSC). Interestingly, significant decoding of layout versus texture was observed in OPA (Natural Open:  $t_8 = 4.00$ ,  $p = 0.016$ ; Manufactured Closed:  $t_8 = 3.27$ ,  $p = 0.046$ ). The similarities between PPA and RSC are consistent with our univariate results, and together with significant decoding in OPA, may speak to a functional dissociation between these regions that are based on differences in processing low-level visual information. However, as caution must be applied both when interpreting null effects with multivoxel data (see Dubois et al., 2015) and when interpreting apparent functional dissociations across cortical regions, we are currently exploring these results in greater detail using a separate paradigm.

#### 3.2.3. Classification Errors

To further investigate the underlying structure of representations across regions and conditions, we next examined, based on the results of the multiclass discriminations, the distribution of classifier guesses via a confusion matrix. In a confusion matrix, each row indicates instances of the actual trial class and each column indicates the trial class predicted by the trained SVM classifier. Thus, the confusion matrix provides not just a visualization of the correct classifications (indicated by classifier responses along the diagonal axis) but also the cases of

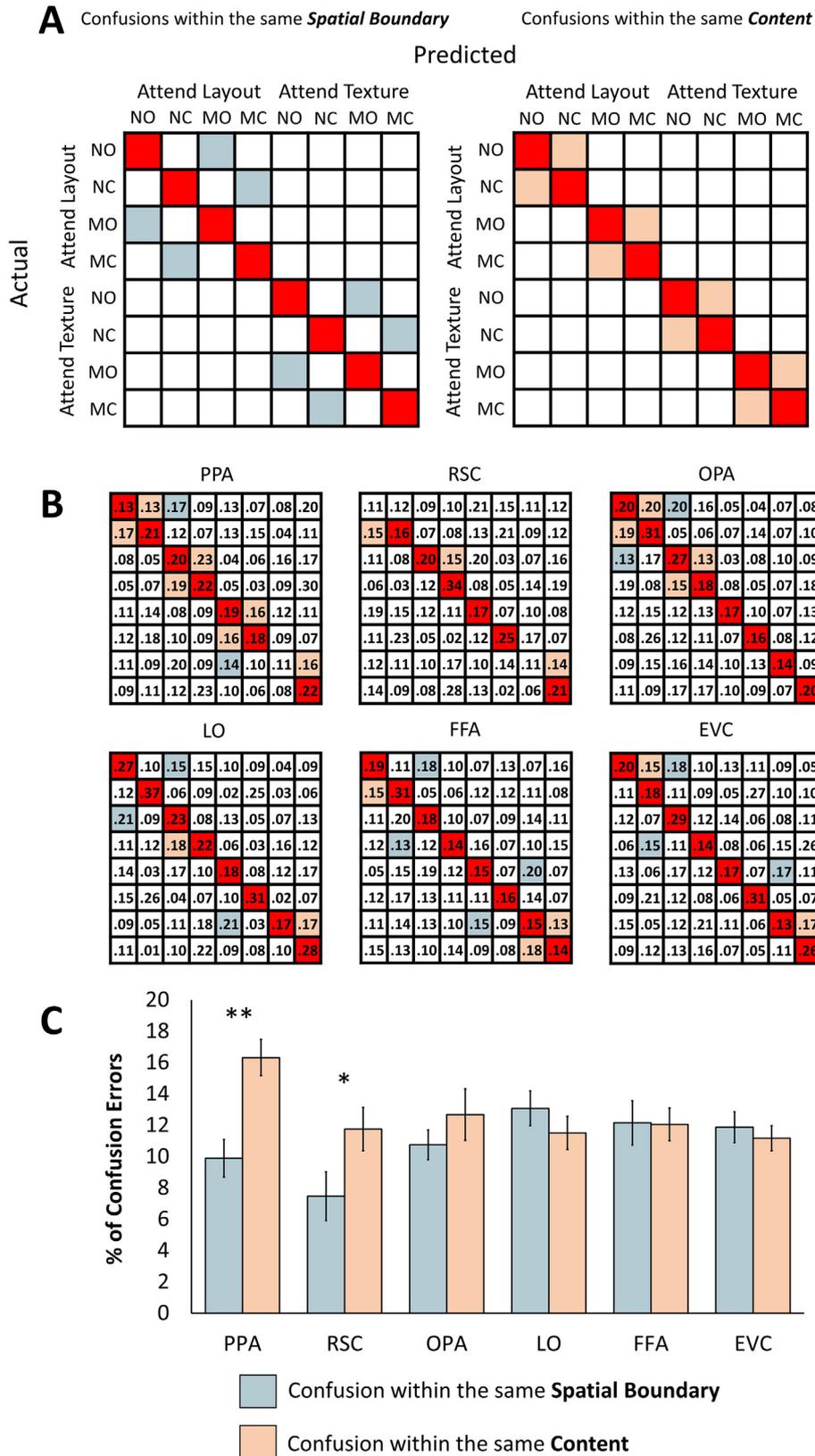


**Fig. 4.** Multivariate results. (A) Classification accuracy (chance = 12.5%; dashed line) of all eight conditions for each ROI. (B) Decoding accuracy (chance = 50%) of scene content with spatial boundary held constant when attending to layout, and (C) spatial boundary with scene content held constant when attending to layout. (D) Decoding accuracy of scene content with spatial boundary held constant when attending to texture, and (E) spatial boundary with scene content held constant when attending to texture. Error bars represent the standard error of the mean. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , + $q \leq 0.05$ .

misclassification (i.e., where the trained classifier “confuses” the actual trial class with that of another class, as indicated by the off-diagonal classifier guesses). The distribution of misclassifications can be informative as it suggests similarity in the patterns of activity across trial types (i.e., two conditions represented similarly are more likely to be misclassified as one another), which is not necessarily evident from the multiclass decoding accuracies alone.

In order to extend upon previous work using a similar analysis (Park et al., 2011), here we determined the types of classification errors made across scene categories while observers attended to individual scene features (texture vs. layout) rather than the image as a whole (as has been done previously). Importantly, this level of analysis allows greater

specification of the types of classifier errors made when attending to different features within a scene and thus greatly improves our understanding of how task demands shape visual representations in scene-selective cortex. If a given region is sensitive to the spatial boundary of a scene when observers attended to either scene layout or texture, then classifier errors in the confusion matrix may be grouped by spatial boundary, regardless of scene content (e.g., open natural scenes may be confused with open manufactured scenes, and closed natural scenes may be confused with closed manufactured scenes, but it is less likely that open and closed scenes would be confused for each other; see hypothetical confusion matrices in Fig. 5A). Conversely, if a given region is sensitive to the content of a scene when observers attended to either



**Fig. 5.** Confusion matrices generated from the multiclass discriminations. (A) Hypothetical confusion errors within the same spatial boundary or within the same content. Note that when decoding is perfect, the confusion matrix will have a diagonal containing values of 1 and the rest of the matrix will be zero. (B) Confusion matrices for each ROI representing classifier errors across conditions. The average classifier response proportions across participants are shown. Shaded squares represent predicted decoding values significantly greater than 12.5% (chance). (C) Confusion errors (collapsed across task) within the same spatial boundary or within the same content across ROIs. Error bars represent the standard error of the mean. N = Natural, M = Manufactured, O = Open, C = Closed, \* $p < 0.05$ , \*\* $p < 0.01$ .

scene layout or texture, then confusion errors may be grouped by scene content, regardless of spatial boundary (e.g., open natural scenes may be confused with closed natural scenes, and open manufactured scenes may be confused with closed manufactured scenes, but it is less likely that natural and manufactured scenes would be confused for each other).

Since distinctive structural features across manufactured and natural environments may be used to accurately discriminate scene content (Walther et al., 2011), a greater number of confusion errors may be observed within the same content, rather than within the same spatial boundary when observers attended to scene layout. For instance, when attending to the layout of a scene, manufactured environments may be confused with each other, but not with natural environments, due to characteristic structural components across manufactured scenes (Oliva and Torralba, 2001), regardless of spatial boundary, which may be of greater importance for distinguishing scenes based on depth. Similarly, spatial boundary may be less important when discriminating texture across scenes. Thus, we hypothesized overall greater confusion errors in PPA within the same content, compared with the same spatial boundary, when observers attended to both the layout and texture of a scene.

Fig. 5B shows the confusion matrices for each ROI, and Fig. 5C shows the types of classification errors made across all ROIs. A repeated-measures ANOVA with factors ROI, scene attribute (spatial boundary vs. content), and task (texture vs. layout) only revealed a significant ROI-by-attribute interaction ( $F_{5,30} = 3.01, p = 0.025$ ), warranting further investigation of differences between spatial boundary and scene content across ROIs. However, since we observed no main effect of task ( $F_{1,6} = 2.60, p = 0.16$ ) and no interactions with this factor (ROI-by-task, task-by-attribute, ROI-by-task-by-attribute; all  $F_s < 2.10$ , all  $p_s > 0.093$ ), we collapsed across task for subsequent analyses. Pairwise comparisons were performed using a two-tailed *t*-test. As expected, confusions within the same content were found to be significantly higher than confusions within the same spatial boundary in PPA ( $t_{11} = 3.29, p = 0.007$ ). Interestingly, RSC showed this same distribution of errors ( $t_{10} = 2.29, p = 0.045$ ), building on both our univariate and multivariate findings showing evidence for similarities in representations between PPA and RSC. In contrast, other regions showed no such differences (all  $t_s < 0.89$ , all  $p_s > 0.40$ ) in the types of classification errors made. Critically, and in line with our univariate and multivariate results, patterns of classification errors in PPA were dissociated from those in LO, FFA and EVC (all  $F_s < 1.53$ , all  $p_s > 0.26$ ), suggesting the patterns of misclassifications observed in PPA were specific to high-level scene processing and were not likely driven by low-level image properties.

### 3.3. Behavioral performance

To ensure attention during the experimental task, observers were asked to compare the identity of two images along a particular relevant feature (i.e., layout or texture). The overall response accuracy was high ( $M = 95.45\% \pm 3.66\%$ , range = 90.67%–97.27%), confirming observers attended to the relevant feature successfully. Behavioral results are reported in Table 1 (we report results for accuracy, and not response time, since we emphasized the former, but not the latter, when describing the task to observers; thus interpreting results for accuracy is more valid in this study). Although near-ceiling performance across all conditions was observed, we did find significant main effects of content ( $F_{1,11} = 8.17, p = 0.016$ ), and task ( $F_{1,11} = 5.46, p = 0.039$ ), but not spatial boundary ( $F_{1,11} = 3.39, p = 0.093$ ). We also observed multiple two-way interactions (content-by-boundary:  $F_{1,11} = 8.08, p = 0.016$ ; content-by-task:  $F_{1,11} = 19.50, p = 0.001$ ; boundary-by-task:  $F_{1,11} = 8.28, p = 0.015$ ). These significant results are likely driven by comparatively lower performance in one condition (see Table 1), but despite this, behavioral response profiles did not match fMRI response profiles in scene-selective cortex (e.g., in manufactured scenes, activation to layout was greater than texture, despite no difference in behavioral

**Table 1**  
Accuracy (percent correct) for each condition.

	Layout	Texture
Manufactured open	95.65 ± 1.11	96.57 ± 0.65
Manufactured closed	96.92 ± 0.57	96.00 ± 1.30
Natural open	97.27 ± 0.58	94.52 ± 0.96
Natural closed	95.96 ± 0.68	90.67 ± 1.20

All values represent mean (percent correct) ± SE.

performance across these conditions). This makes it unlikely that task difficulty directly contributed to the observed neural activation patterns in scene-selective cortex. This is consistent with previous findings showing that modulation of activity in PPA was dissociable from manipulations of task difficulty (Xu et al., 2007).

## 4. Discussion

It has previously been proposed that the diagnostic structure and meaning of a visual scene is characterized by the boundaries and content of a space and is captured by a collection of global image features, such as texture and layout (Oliva and Torralba, 2006). Here, we examined the underlying neural representations of global scene texture and structural layout to explore the contributions of these features in scene processing, testing the hypothesis that the diagnostic relevance of these features in their respective scenes would flexibly modulate activity in scene-selective cortex. We present novel evidence demonstrating that, relative to layout, activity in scene-selective cortex showed equal sensitivity to texture in natural scenes, but less sensitivity to texture in manufactured scenes. These findings indicate activity in scene-selective cortex is not only modulated by multiple scene features but may be scene-specific and dependent on the relevance of various features within a scene. Critically, this pattern of univariate activation differed markedly from that observed in early visual cortex, and areas selective to face and object processing, demonstrating these findings are unique to scene-selective cortex, and are not simply a result of differences in low-level image properties. In line with these findings, multivoxel pattern analysis revealed that the encoding of high-level scene attributes varies according to scene category and is influenced by task context, suggesting dynamic and flexible scene representations are formed from an interaction between multiple scene properties and the task goals of the observer.

### 4.1. Feature diagnosticity and task context in scene recognition

In support of previous theories of a primary role for PPA in spatial encoding (Epstein and Kanwisher, 1998; Epstein et al., 2003; Park et al., 2011; Kravitz et al., 2011), the data presented here demonstrate overall greater sensitivity to the encoding of structural layout, relative to texture, across scene-selective cortex. Typically, edge-based structural information is sufficient for observers to distinguish between scenes (Biederman and Ju, 1988; Delorme et al., 2000; Walther et al., 2011). In contrast, our results indicate that when structural information is less informative (e.g., in natural environments), processing in scene-selective cortex may rely on other diagnostic features which uniquely inform scene representations. Indeed, Steeves and colleagues (2004) have shown that, unlike in healthy observers, PPA activation in an individual with profound visual form agnosia (i.e., impairments in processing structure) was modulated by the presence of appropriate scene color when color was useful for accurate scene identification. Neural activity in scene-selective regions within the ventral visual cortex may therefore also reflect the relative high-level contributions of various scene features beyond spatial layout which contribute to the formation of scene identity.

In a similar vein, Harel et al. (2014) have suggested top-down signals produced by behavioral goals and observer intent directly impact visual

object representations within the ventral visual pathway, supporting the view that cortical activity reflects not only the physical properties of a stimulus, but also the internal state of the observer. In the present study, we provide evidence consistent with this research by showing that task context (i.e., attended feature) influences neural activity in scene-selective cortex. Here, global texture cues may provide meaningful information for discriminating scenes containing a high degree of physical similarity, such as desert landscapes (see Fig. 1B), and thus attending to these features may selectively modulate activity in scene-selective cortex accordingly. Furthermore, in addition to providing meaningful high-level identity information, global texture cues have been shown to inform the processing of spatial depth and contour (Torralba and Oliva, 2003), suggesting a convergent and complementary relationship between texture and spatial structure.

#### 4.2. Decoding of high-level scene attributes

Moving beyond the question of whether a brain region is sensitive to certain visual features, we next aimed to explore discriminations between various high-level scene attributes in ventral visual cortex. Consistent with previous research suggesting PPA represents both content and spatial boundary information (Park et al., 2011), our multivariate analysis revealed that regions in scene-selective cortex discriminated between both of these high-level scene attributes. Having confirmed previous findings, we next employed novel analyses to determine whether the decoding of high-level attributes varies according to task context and scene category. Investigation of within-category discriminations revealed significant decoding of spatial boundary and layout when attending to both texture and layout, suggesting that, together with layout, texture may form an important basis for defining scene identity. More specifically, this analysis revealed results consistent with our univariate findings: significant decoding of a spatial scene feature (i.e., boundary) only occurred in manufactured, but not natural scenes (in PPA and RSC), when observers attended to the layout of a scene. The decoding of spatial boundary when observers attended to texture, however, differed markedly from when observers attended to layout. Critically, these findings show for the first time that task context directly impacts the representations of high-level scene attributes and underscores the notion that scene attributes and task demands may modulate activity in scene-selective cortex to varying degrees across different scene categories. Together with our univariate results, these novel findings suggest the importance of scene features may be scene-specific and task-dependent, rather than of equal importance across different scene categories.

In contrast to the above results, both spatial boundary and content could be decoded when attending scene layout regardless of scene type in areas of early visual cortex, consistent with previous research (Park et al., 2011). Real-world scene categories contain a high degree of statistical regularity reflected in distinct low-level features and global spatial frequency distributions across both scene categories and boundaries (Oliva and Torralba, 2001; Torralba and Oliva, 2003). This leaves open the possibility that scenes could be discriminated based on low-level attributes alone. Although it is highly likely that low-level features and statistical regularities across scene categories contribute to the activation patterns associated with scenes in PPA, these neural patterns are unlikely to be driven purely by such features. Indeed, in addition to the functional dissociations between PPA and EVC reported here (Figs. 3, 4C, D, and 5C), previous findings suggest EVC plays a less direct role in humans' ability to categorize real-world scenes compared with PPA (Walther et al., 2009).

#### 4.3. Misclassifications across scene categories

Contrary to previous work (Park et al., 2011), our analysis of the distribution of classification errors across scene conditions via the confusion matrix revealed greater clustering of errors within the same

content in PPA and RSC, compared with the same spatial boundary. Moreover, our results further differ from this previous work in the finding that the pattern of misclassifications in PPA and RSC were dissociable from the pattern observed in early visual cortex. These findings suggest that the structure of encoding observed in PPA extends beyond the low-level properties of a scene and reflects the high-level representations of particular features within a scene. For instance, misclassifications clustered within the same content may be indicative of an increased reliance on cues provided by relatively stable structural differences across scene categories. These cues may support the discrimination of features such as layout and texture, whereas spatial boundary attributes may be less relevant for distinguishing these features, and instead may be more relevant for determining spatial depth and routes for navigation.

How can we reconcile our findings with those from previous work? Our task was unique as we examined attention to specific features within a scene (i.e., layout vs. texture), and as such, we propose the differences between our findings and those from other researchers stem from the task demands and goals of the observer (Harel et al., 2014), the diagnostic relevance of different features across scene categories and the differential allocation of attention to these features, and differences across the stimuli themselves. For instance, in order to control for interference effects of salient objects (Davenport and Potter, 2004; Joubert et al., 2007), our scenes were selected to be devoid of foreground objects, which would likely explain differences in the representations of information in object-selective cortex when compared with previous work, which found greater confusions of content in LO (Park et al., 2011). Indeed, across our analyses, activity patterns in LO were dissociated from those in scene-selective cortex and did not show differences in confusion errors distributed across high-level scene attributes.

Finally, to the best of our knowledge, previous work has not revealed how manipulating attention to a particular scene feature has differential effects on activity in scene-selective cortex, depending on the type of scene being viewed. Thus, these current results emphasize the fundamental importance of examining the modulation of activity in scene-selective cortex as a function of attending to various scene-specific properties and task-based goals. In the real world, we are not merely passive observers for the purposes of perception but use attention to filter the influx of visual information in accordance with our goals. That is, we engage in scene processing for the purposes of interacting with our environment, and based on the nature of the surrounding environment, different visual features will become more or less important.

#### 4.4. Functional representations across the scene-processing network

The results of the present study have demonstrated both similarities and differences in functional representations of scenes across the broader scene-processing network (PPA, RSC, and OPA). These regions exhibited similar patterns of univariate activation when observers attended to either scene texture or layout, suggesting reliable and distinguishable representations of diagnostic scene features across the scene-processing network compared with non-scene-selective regions (e.g., FFA, EVC). Conversely, the representation of high-level scene attributes (content and spatial boundary) differed across these regions. Across our analyses, similar patterns of activation were observed in PPA and RSC, yet these patterns were often dissociated from those observed in OPA (see Figs. 3, 4C, and 5C). Damage to PPA and RSC has been associated with deficits in the simple visual identification of scenes or landmarks (Aguirre and D'Esposito, 1999; Mendez and Cherrier, 2003), and with difficulty in an individual's ability to use landmarks to orient themselves in order to navigate through an environment (Takahashi et al., 1997), respectively. While recent work has demonstrated a causal involvement for OPA in scene perception (Dilks et al., 2013), less is known about how different scene properties are represented within this region.

The functional differences we observe across PPA, RSC, and OPA may be related to differences in how task context and the importance of perceptual features shape representations across areas of scene-selective cortex. In other words, the goals of the observer may differentially affect activation across PPA, RSC, and OPA, by interacting with the type of perceptual information processed within these regions (e.g., both spatial and non-spatial visual information in PPA, but more weighting towards spatial information in OPA; Cant and Xu, 2012; In Press). An important future question concerns whether OPA represents high-level information consistent with PPA and RSC, or more basic perceptual information which may complement representations in PPA and RSC. Here, we present a broad array of evidence clearly demonstrating dissociations between OPA and both PPA and RSC, but additional research is needed to clarify the nature of these dissociations.

## 5. Conclusion

To date, there has been considerable debate surrounding the structure of representations in scene-selective cortex. While support for spatial encoding has been widespread, neural evidence for how texture is processed within the context of a scene has received little study, despite computational and psychophysical evidence supporting a meaningful role for global texture in scene discrimination. Here, we demonstrate that scene-selective cortex represents multiple visual features, such as layout and texture, that the representations of these features are shaped by perceived scene category, and that these results are dissociated from activity in early visual cortex and areas non-selective to scene processing. Furthermore, differences in the multivariate patterns of activation observed across regions of scene-selective cortex suggest differentiation in the representations of visual features and scene attributes across the broader scene-processing network. The present study therefore ties together multiple factors (high-level scene attributes, task context, and individual visual features), which were not jointly investigated previously, to highlight the fact that scene perception and recognition, and visual processing more generally, are performed to serve a particular goal, and it is this goal which determines how informative a particular visual feature is in a particular environment.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

This study was supported by a Natural Sciences and Engineering Research Council (NSERC) grant (435647-13) and a Connaught New Researcher Award to J.S.C., and an NSERC grant (216203-13) and Canadian Institutes of Health Research (CIHR) grant (106436) to S.F. J.P.G was supported by a CIHR postdoctoral fellowship.

## Acknowledgments

The authors thank Dirk Bernhardt-Walther for helpful comments and discussion and Nancy J. Lobaugh for help with MRI acquisition.

## References

Aguirre, G.K., D'Esposito, M., 1999. Topographical disorientation: a synthesis and taxonomy. *Brain* 122 (9), 1613–1628.

Bar, M., Aminoff, E., Schacter, D.L., 2008. Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J. Neurosci.* 28 (34), 8539–8544.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 289–300.

Biederman, I., Ju, G., 1988. Surface versus edge-based determinants of visual recognition. *Cogn. Psychol.* 20 (1), 38–64.

Cant, J.S., Goodale, M.A., 2007. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17 (3), 713–731.

Cant, J.S., Goodale, M.A., 2011. Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. *J. Neurosci.* 31 (22), 8248–8258.

Cant, J.S., Xu, Y., 2012. Object ensemble processing in human anterior–medial ventral visual cortex. *J. Neurosci.* 32 (22), 7685–7700.

Cant, J.S., Xu, Y., 2015. The Impact of density and ratio on object-ensemble representation in human anterior–medial ventral visual cortex. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bhu145> (In Press).

Castelhano, M.S., Henderson, J.M., 2008. The influence of color on the perception of scene gist. *J. Exp. Psychol. Hum. Percept. Perform.* 34 (3), 660.

Davenport, J.L., Potter, M.C., 2004. Scene consistency in object and background perception. *Psychol. Sci.* 15 (8), 559–564.

Delorme, A., Richard, G., Fabre-Thorpe, M., 2000. Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vis. Res.* 40 (16), 2187–2200.

Dilks, D.D., Julian, J.B., Kubilius, J., Spelke, E.S., Kanwisher, N., 2011. Mirror-image sensitivity and invariance in object and scene processing pathways. *J. Neurosci.* 31 (31), 11305–11312.

Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* 33 (4), 1331–1336.

Dubois, J., de Berker, A.O., Tsao, D.Y., 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci.* 35 (6), 2791–2802.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. Second ed Wiley Interscience (Chapter 3).

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392 (6676), 598–601.

Epstein, R.A., Higgins, J.S., 2007. Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb. Cortex* 17, 1680–1693.

Epstein, R., Graham, K.S., Downing, P.E., 2003. Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37 (5), 865–876.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2 (4), 189–210.

Gallivan, J.P., McLean, D.A., Flanagan, J.R., Culham, J.C., 2013. Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas. *J. Neurosci.* 33 (5), 1991–2008.

Gallivan, J.P., Cant, J.S., Goodale, M.A., Flanagan, J.R., 2014. Representation of object weight in human ventral visual cortex. *Curr. Biol.* 24 (16), 1866–1873.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., Rossion, B., 2005. Diagnostic colours contribute to the early stages of scene categorization: behavioural and neurophysiological evidence. *Vis. Cogn.* 12 (6), 878–892.

Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3 (8), 837–843.

Harel, A., Kravitz, D.J., Baker, C.I., 2014. Task context impacts visual object processing differentially across the cortex. *Proc. Natl. Acad. Sci.* 111 (10), E962–E971.

Hsu, C.W., Lin, C.J., 2002. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* 13 (2), 415–425.

Joubert, O.R., Rousset, G.A., Fize, D., Fabre-Thorpe, M., 2007. Processing scene context: fast categorization and object interference. *Vis. Res.* 47 (26), 3286–3297.

Kravitz, D.J., Peng, C.S., Baker, C.I., 2011. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J. Neurosci.* 31 (20), 7322–7333.

MacEvoy, S.P., Epstein, R.A., 2011. Constructing scenes from objects in human occipitotemporal cortex. *Nat. Neurosci.* 14, 1323–1331.

Mendez, M.F., Cherrier, M.M., 2003. Agnosia for scenes in topographagnosia. *Neuropsychologia* 41 (10), 1387–1395.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53 (1), 103–118.

Oliva, A., Schyns, P.G., 1997. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* 34, 72–107.

Oliva, A., Schyns, P.G., 2000. Diagnostic colors mediate scene recognition. *Cogn. Psychol.* 41 (2), 176–210.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42 (3), 145–175.

Oliva, A., Torralba, A., 2006. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* 155, 23–36.

Park, S., Brady, T.F., Greene, M.R., Oliva, A., 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J. Neurosci.* 31 (4), 1333–1340.

Peuskens, H., Claeys, K., Todd, J., Norman, J., Hecke, P., Orban, G., 2004. Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J. Cogn. Neurosci.* 16 (4), 665–682.

Rao, A.R., Lohse, G.L., 1993. Identifying high level features of texture perception. *CVGIP: Graph. Model. Image Process.* 55 (3), 218–233.

Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. *Neuroimage* 30 (4), 1088–1096.

- Schyns, P.G., Oliva, A., 1994. From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5 (4), 195–200.
- Steeves, J.K., Humphrey, G.K., Culham, J.C., Menon, R.S., Milner, A.D., Goodale, M.A., 2004. Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J. Cogn. Neurosci.* 16 (6), 955–965.
- Takahashi, N., Kawamura, M., Shiota, J., Kasahata, N., Hirayama, K., 1997. Pure topographic disorientation due to right retrosplenial lesion. *Neurology* 49, 464–469.
- Talairach, J., Tournoux, P., 1988. *Co-planar stereotaxic atlas of the human brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Thieme, Stuttgart, Germany.
- Torralba, A., Oliva, A., 2003. Statistics of natural image categories. *Netw. Comput. Neural Syst.* 14 (3), 391–412.
- Walther, D.B., Shen, D., 2014. Nonaccidental properties underlie human categorization of complex natural scenes. *Psychol. Sci.* 0956797613512662.
- Walther, D.B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 29 (34), 10573–10581.
- Walther, D.B., Chai, B., Caddigan, E., Beck, D.M., Fei-Fei, L., 2011. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc. Natl. Acad. Sci.* 108 (23), 9661–9666.
- Xu, Y., 2010. The neural fate of task-irrelevant features in object-based processing. *J. Neurosci.* 30 (42), 14020–14028.
- Xu, Y., Chun, M.M., 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440 (7080), 91–95.
- Xu, Y., Turk-Browne, N.B., Chun, M.M., 2007. Dissociating task performance from fMRI repetition attenuation in ventral visual cortex. *J. Neurosci.* 27 (22), 5981–5985.